Appendix A: Methodology

A.1 Sample design

This section provides a brief overview of the sample design used for the National Diet and Nutrition Survey Rolling Programme (NDNS RP) Years 9 to 12 (2016/17 to 2018/19) from which eligible participants (1) were approached to take part in this study.

The NDNS RP has run continuously since 2008 and is a cross-sectional survey, with a sampling frame designed to draw a sample of people representative of the UK population aged 1.5 years and over each survey year. The sample for each NDNS RP year is drawn from the 'small users' sub-file of the Post Office's Postcode Address File (PAF), a computer list of all the addresses (delivery points) which receive fewer than 25 articles of mail a day. In order to improve cost effectiveness, addresses are clustered into Primary Sampling Units (PSUs), small geographical areas, based on postcode sectors, randomly selected from across the UK. A list of addresses is randomly selected from each sampled PSU. In order to achieve equal numbers of adults and children in the sample, the selected addresses are randomly allocated to 1 of 2 groups to determine whether adult(s) (aged 19 years or over) and children (aged 1.5 to 18 years) at "basic addresses", or children only at "child boost addresses", are selected for NDNS RP interview. At each address, the interviewer enumerates the number of households and, in cases where there are 2 or more, randomly selects one household for the NDNS RP. Prior to Year 12, the interviewer randomly selected up to one adult and one child from each household selected for invitation to take part in the NDNS RP. In Year 12, the interviewer selected up to 2 adults and one child (at basic addresses) and up to 2 children (at child boost addresses) to take part (2). Further information on the sample design for NDNS RP Years 9, 10 and 11 can be found in appendix B of the NDNS RP Years 9 to 11 report (3).

The sample drawn for this study comprised adults and children who had previously taken part in Years 9 to 12 (2016/2017 to 2019/2020) of the NDNS RP and who had consented to being re-contacted regarding future research. A total of 69 NDNS RP Year 12 participants were not included in the sample for this study as their NDNS RP data was still being processed and cleaned at the time the sample was drawn. Of those who participated in NDNS RP Years 9 to 12, 84% were sent a letter inviting them to take part in this study. This approach was intended to generate a sample size large enough to allow for data to be analysed by age and sex while keeping the time period between data collections as short as possible so that participant contact details were less likely to have changed and to minimise change factors between NDNS RP participation and this study.

A.2 Contacting participants

The study fieldwork period ran from 10 August to 31 October 2020. Fieldwork was issued in 3 batches, on 10 August 2020 and then 9 and 25 days later. Invitation letters were posted on different days of the week to balance out the days of the week on which participants would likely be invited to complete their dietary recalls to maximise coverage of days of the week in the dietary dataset. The first 2 batches of invitation letters were sent to participants from NDNS RP Years 9 to 11 (letters were sent on 10 August and 19 August 2020). Batch 3 contained participants from NDNS RP Year 12, with letters sent on 4 September 2020.

Seven days after dispatch of the invitation letters, individuals who had not completed the web questionnaire were contacted by the NatCen Telephone Unit (TU) to encourage participation in the study. The TU interviewers directed individuals to the invitation letter, explained the study to them and responded to any questions. The TU interviewer provided the questionnaire URL and individual's unique access code or provided assistance to complete the questionnaire if required. Using the contact details provided in their original NDNS RP interview, NDNS RP Year 12 participants also received an email or text re-invitation message (appendix E), 7 days after dispatch of their invitation letter prompting them to take part in the study. These participants then received their TU phone call a few weeks after their initial invitation letter and the email or text re-invitation message, if they had not completed the web questionnaire. NDNS RP Years 9 to 11 (2016/17 to 2018/19) participants were not contacted by email or text due to concerns about whether the contact information collected during their original interview would be up to date.

A.3 Dietary assessment

A.3.1 Introduction

For Years 1 to 11 (2008/09 to 2018/19) of the NDNS RP, dietary data was collected over 4 consecutive days using a paper food diary with estimated weights. From Year 12 (2019 to 2020), the NDNS RP moved to collecting 4 non-consecutive 24-hour dietary recalls using Intake24, an online dietary data collection tool in place of the paper diary. This change was made to provide a more automated data collection method, compatible with new technologies and with potential to improve data quality and reduce costs. The introduction of Intake24 in the NDNS RP represents a significant transition for the survey. A formal evaluation is being undertaken within the NDNS RP to assess the impact of the method change on the survey data. The report on the first stage of this evaluation has been published and is based on all available data from Year 12 (October 2019 to March 2020). It provides a set of interim comparison analyses to aid interpretation of the findings from this study (4).

A.3.2 Food diary

Participants who originally took part in the NDNS RP in Years 9 to 11 (2016/17 to 2018/19) were asked to keep a record in their paper food diary of everything eaten or drunk during the full 24-hour period of each recording day. Food portion photographs were provided for a small number of frequently consumed foods but the majority of portion sizes were recorded by participants in household measures (for example, 2 dessert spoons of baked beans, one Kit Kat (2 fingers) or, for packaged foods, the weight indicated on the packet. Foods and portions were then entered by trained coders into the dietary assessment system DINO (Diet In Nutrients Out) (5). The linked food composition data came from respective years of the Public Health England (PHE) NDNS Nutrient Databank (NDB).

Details of the food diary methodology can be found in appendix A of the NDNS RP Years 9 to 11 report (3).

A.3.3 Intake24

For those participants who originally took part in the NDNS RP in Year 12 (2019 to 2020) their previous dietary data was collected using Intake24. Dietary data for this study (August to October 2020) was also collected using Intake24.

Intake24 is an online automated self-administered 24-hour dietary recall tool based on the multiple pass method and has been validated against doubly labelled water (DLW) in adults (6,7,8).

Participants completing Intake24 are asked to record everything they ate and drank the previous day (midnight to midnight). Participants are guided through webpages in the following order:

1. Meal-based quick list (captures time, and lists foods)
2. Detail pass (foods selected, portion details provided)
3. Meal gap review
4. Last chance review

The tool displays pre-defined meal and snack times (for example, breakfast at 08:00) which are displayed in chronological order. However, the user can choose to complete entries in any order they wish, change the time and can delete or add more eating occasions. The multiple pass method means that foods and drinks are first entered as free text in the 'quick list' phase of the dietary recall. This is to collect a list of foods and drinks consumed during the previous day. The participant is then taken through the 'quick list' and for each item reported is asked to select from a list of matching food descriptions linked to a food code within Intake24. Participants can add or delete foods in a meal. Once a food is selected, participants are presented with portion

size estimation options appropriate to the particular food or drink. The majority of foods in Intake24 have a range of portion-size photo images that the participant can review and select for their portion size. If there is no photo, there is the option to report portion size as household measures such as individual items, different spoon sizes, or small, medium or large servings. Some foods offer both photos and household measures.

Intake24 includes a custom spell checker which works to correct both phonetic misspellings and typing errors in the free text. The system also prompts participants for items commonly consumed together (for example, milk in tea or coffee), and prompts for missing foods and drinks if there are long time gaps between reported eating occasions or where only a few foods have been reported. A demo version of the Intake24 system is available at Intake24 (9).

Participants are asked to select, as far as possible, foods from the food lists provided in Intake24. If they cannot find an exact match for the food consumed the tool prompts them to choose the closest matching item where possible. If the participant still cannot find a suitable match for their food they can report this as a missing food. The missing food function asks the participant to provide key information about the food for coding at a later stage by the research team based in Cambridge. After reporting each eating occasion participants are asked to record where they had bought or obtained most of the food for that meal or snack (see appendix F). At the end of each dietary recall, questions prompt participants to record if their intake was typical for that day (and if not, the reason why), any special diet, the type of oil generally used for cooking (if any) and details of any dietary supplements taken. For this study participants were also asked after each dietary recall if they or someone in their household was self-isolating or shielding (10).

For children aged 10 years or younger, a parent or carer was asked to complete the dietary recall with input from the child as appropriate. Children aged 11 years and older could complete the recall themselves with details confirmed with others where necessary. If the dietary recall was completed by a proxy there may have been occasions when they were unsure of the details of a particular food or an entire meal (such as for a child attending school). At the end of each dietary recall there was a question asking if the recall had been completed by a proxy (that is on the selected participant`s behalf) and, if yes, whether there were any difficulties. Any gaps in the recall could be noted here and taken into account during the data checking process.

At the end of the fourth completed dietary recall, Intake24 provided participants with the option to view on-screen dietary feedback. The feedback provided an averaged daily personal intake for key foods and nutrients compiled from the participant's 4 dietary recalls combined. See appendix E for screenshots of the dietary feedback. Participants who completed fewer than 4 recalls were not provided with feedback.

A.3.3.1 Intake24 assistance
For this study, participants who did not have internet access or who did not feel confident completing the dietary recalls independently could request assistance for their dietary data

collection. Assistance with completing dietary recalls was provided by the Cambridge team who carried out assisted dietary recalls over the telephone or via Zoom video conferencing. Participants opting for telephone dietary recall assistance were sent a hard-copy food photograph atlas in advance to aid the estimation of portion sizes during completion of the dietary recalls. The researcher read out the instructions and prompts in Intake24 with the participant or parent/carer providing the information on their food and drink consumption for the previous day which was then entered into Intake24 by the researcher. Participants being assisted over the telephone were asked to refer to the food photo atlas to select their portion size or were asked to describe the amount in household measures. At the end of the dietary recall, the researcher arranged an appointment for the next dietary recall if assistance was still required.

If all 4 dietary recalls were completed with assistance, the participant was sent a hard copy of their dietary feedback via email or post if they wished.

A.3.3.2 Intake24 dietary data processing

As with the coding of the paper diaries in the NDNS RP, all food codes within Intake24 are assigned nutrient composition data aligned with the PHE NDB for the respective year (see section A.3.4). Food and portion size codes are automatically assigned within Intake24 allowing the system to generate nutrient output at the individual food level. The raw Intake24 output was imported into a bespoke database to facilitate data checks and to assign foods reported as missing to an appropriate food code and portion size, using the original free text search term and missing food details provided by the participant.

A.3.3.3 Intake24 quality control

During data collection, monitoring checks were carried out on the dietary recalls by the Cambridge team which were:

- fewer than 10 food items in a recall (excluding dietary supplements and associated foods for example, milk with tea)
- 3 or fewer eating occasions in a recall (this includes occasions when a participant reports consuming only a drink without food)
- completion time of under 3 minutes
- total energy intake less than 400kcal or more than 4,000kcal and the participant had not stated that they consumed 'less than usual', 'more than usual', or that they were on a weight gain or weight loss diet
- participant had written multiple food items in their search term, for example, 'toast, cereal, yoghurt' but only one food item was coded
- inconsistencies between the search term and the food code selected (for example, searched for chicken stir-fry but selected prawn stir-fry)
- 'orphan' foods (a reported food that appeared to have been eaten on its own for example, beef steak when it would commonly be eaten with other foods such as chips, potatoes or salad)

These checks aimed to monitor the performance of Intake24, to identify potential issues and improvements for tool functionality and usability, and also to identify and resolve any issues in the study dietary datasets.

At the end of data collection, and when all missing foods had been coded, boxplots were generated by NDNS standard sex and age groups (1.5 to 3 years, 4 to 10 years, 11 to 18 years and 19 years and over) to review portion sizes and to identify any extreme outliers within each food group (11). Extreme outliers were identified from the boxplots as individual data points separate from the box and whiskers since they were more than 1.5 x IQR (Inter-quartile range: 75th percentile to 25th percentile) from the nearest quartile for that intake (either the 25th or 75th percentile). These were examined on a case-by-case basis and reviewed in the context of the participant's overall consumption. Portion sizes which were considered to be implausible, and potentially the result of errors in portion size selection, were adjusted. Adjustments were carried out in the bespoke dietary database by changing the portion code at the individual recall level (12).

Finally, boxplots were generated by age group to identify any infeasible or extreme energy and nutrient values. As with portions, extreme outliers were looked at on a case-by-case basis. Extreme intakes that were considered to be the result of errors in portion size or food composition in the NDB were adjusted, otherwise values were left in the dataset as they were assumed to reflect consumption by participants.

As a result of the extreme outlier checks, 10.8% of food items in this study were adjusted. No dietary recalls were excluded from the analysis as a result of the dietary data checks.

A.3.4 NDNS nutrient databank (NDB)

A.3.4.1 Updates to the NDB

Dietary data for all years of the NDNS RP and for this study is linked to food composition data for respective years of the PHE NDNS NDB. Each year a programme of updates and revisions are made to the NDB so that the databank is up-to-date and, as far as possible, reflects the nutrient composition of the food supply for each survey year reported. Updating of the NDB may include the addition of new foods as well as revision of nutrient composition of existing foods, either at food group level following a programme of reanalysis, or to take account of reformulation reported by manufacturers and changes in fortification practices. Therefore the same foods may have a different composition for some nutrients in one year of the NDNS RP compared to another year. Table A.1 lists the updates to the NDB that were applied to the NDNS RP Year 12 (2019 to 2020) and this study. Further details on the NDB can be found in the Years 9 to 11 report (3).

Table A.1 Updates to the NDB applicable to NDNS RP Year 12 and this study

Analytical data updates	Reviews of manufacturers' or label data
None	Bacon and ham[a]
	Baked beans[a]
	Bread and rolls[a]
	Breakfast cereals
	Butter and fat spreads
	Cheddar cheese[a]
	Cook in and pasta sauces[a]
	Crisps
	Pizza
	Sausages
	Soft drinks
	Soup[a]
	Sweet biscuits
	Yoghurt, fromage frais and other dairy desserts

[a] Only sodium values updated

A.3.4.2 Rationalisation of the NDNS NDB for Intake24

Before Intake 24 was introduced into the NDNS RP in Year 12, a major review and update of the food lists within Intake24 was undertaken so that they were simple and usable for participants while reflecting as far as possible the heterogeneity of the food supply in the UK. All foods had to be linked to appropriate portion options and comprehensive nutrient composition values in the most up-to-date NDB. This review process also provided an opportunity to rationalise the NDB to allow more efficient management of updates in future years of NDNS RP. Details of this review and rationalisation can be found in the evaluation report (4).

A.4 Physical activity assessment

Participants aged 16 years and over who had completed all 4 dietary recalls were invited to complete a physical activity assessment using the Recent Physical Activity Questionnaire (RPAQ) (13). RPAQ was developed by the MRC Epidemiology Unit, Cambridge and has been used in each year of the NDNS RP with some minor adaptations and variations in questions between years.

The RPAQ is a self-completion questionnaire designed to assess an individual's physical activity over the previous 4 weeks. It contains questions about physical activity in 4 domains: at home, at work, during commuting and during leisure time. Questions are closed rather than open-ended to facilitate completion, large-scale data entry and analysis.

The RPAQ has been validated against DLW and individually calibrated heart rate and movement sensing to assess physical activity energy expenditure (PAEE) in adults (14,15).

From NDNS RP Year 12 (2019 to 2020) the RPAQ moved from a paper questionnaire used in previous NDNS RP years to an online version. The online version of RPAQ was used for this study. Minor wording and question changes were made between the paper and online versions. Separate questions on television watching and computer use were included in the paper version whereas in the online version a single question was asked about overall screen time as this reflects current screen use habits. During NDNS RP Year 12 participants completed the online version during the interviewer visit, whereas for this study, participants were sent by text or email (after completing their fourth recall) a unique web link to self-complete the RPAQ online with no interviewer present.

Total PAEE was calculated by summing up PAEE for all activities across all domains. Total PAEE and PAEE in each domain (at home, at work, during commuting and during leisure time) were expressed in kJ per kg per day where one metabolic equivalent (MET) equates to 71.2 J per min per kg (3.5 ml O_2 per min per kg).

To compute PAEE from the RPAQ, reported time spent on activities was multiplied by the metabolic cost of each activity (in metabolic equivalents, METs) obtained from the physical activity compendium (16,17) minus one MET for resting metabolic rate, to provide activity-specific PAEE estimates (14). Modifications to this scoring regime were made for occupational activity which was quantified according to the approach outlined by Golubic and others (15) and derived from a cohort of 12,435 UK adults in the Fenland Study who had completed the RPAQ concurrently with objective assessment of PAEE, estimated from individually calibrated combined heart rate and movement sensing (18). The average intensity estimates for reported work duration were 1.54 METs for sedentary occupations, 1.74 METs for standing occupations, 1.93 METs for manual work, and 2.20 METs for heavy manual work. If total reported time spent across all activities was greater than 18 hours per day (assuming 6 hours sleep), all reported durations of activity were scaled back to 18 hours for that person (14).

None of the changes between the paper version used in NDNS RP Years 9 to 11 (2016/17 to 2018/19) and the online version used in NDNS RP Year 12 (2019 to 2020) and this study affect PAEE calculation.

A.5 Web questionnaire data processing

Following the completion of fieldwork, the web questionnaire data was processed and cleaned and, where relevant (for example month or year of birth), details were reconciled between data collection for this study and the participant's original NDNS RP interview.

Participants were asked in the web questionnaire whether their address had changed since their original NDNS RP participation. Three households had disbanded since their NDNS participation and not all NDNS RP participants in these households still lived together. These participants were treated as living in separate households.

The questionnaire for this study, which each NDNS RP participant was invited to complete, included some questions about the individual and some questions about the household as a whole. In the NDNS RP, household-level questions are asked of only one person in a household (usually the Household Reference Person (HRP) or Main Food Provider (MFP)) on behalf of the household. Individual-level questions would then be asked of each individual participant. In this study however, each participant was asked the same set of questions, including questions about their household, as it was not possible with the remote study design to identify and select an HRP nor anticipate which (if any) of the invited participants would take part.

Therefore, household-level data collected from more than one individual in multiple-participant households was harmonised to allow for reporting at the household level. For single response-code questions, the responses of the oldest participant were used to represent the household. For household-level multicode (that is, 'select all that apply') questions, all response codes selected by household participants were included in the household response, such that if any participant in the household selected an answer option, that response was included for the household as a whole. This process resulted in 14 households reporting conflicting information, for example, one participant reporting the household 'cooked at home more' while another participant reporting the household 'cooked at home less'. In these scenarios, the responses cancelled each other out, resulting in both responses being set as 'not selected' for the household.

Participants aged 11 to 15 years were asked to complete the web questionnaire themselves but were prompted to ask their parent or carer for support answering the household-level data. Due to the concern that some of these household-level questions may have been answered by children without parent or carer support, and given the nature of these questions, these data were excluded from the analysis and reporting of household-level questions (participants aged 11 to 15 years were still included in individual-level reporting). Household-level data from these participants were therefore excluded prior to the harmonisation process described above. In total, 101 households were excluded from household-level analysis and reporting as they contained participants aged 11 to 15 years only (and no adult participant).

Household-level data from participants aged 2 to 10 years was included as the parent or carer was directed to complete the whole questionnaire for these young participants. Three households contained participants aged 2 to 10 years and 11 to 15 years (and no adult participant). The household-level response of the younger child was included in the analysis as the questionnaire was completed by their parent or carer.

A.6 Weighting

Weights were required for the data to remove any bias in the observed results due to differences in selection probabilities and non-response. Different households and individuals had different chances of being selected to take part; these vary by country of residence, household size, and survey year. Boosted samples were collected in Wales (Year 9) and Northern Ireland (Years 9 and 11) to permit analysis at country level. Cases collected as part of the boosted sample were weighted down in the dataset to reflect the UK population. In addition, there were a number of points at which respondents could drop out of the survey and cause non-response bias.

Four sets of weights were required which are:

- a set of weights for the 1,046 individuals who had completed the web questionnaire
- a set of weights for the 930 individuals who had completed at least one dietary recall
- a set of weights for the 373 adults who had completed the physical activity module (RPAQ)
- a set of household-level weights for the 801 households with a completed questionnaire for at least one household member not aged 11 to 15 years (19)

The weights were generated in a number of stages. These are described in detail in the following sections but in summary, these stages were:

- the starting point was to take the weights from NDNS RP Years 9 to 12, then adjust these for the initial household and individual selection probabilities and non-response to the NDNS RP
- an adjustment was then made to account for differences in the likelihood that NDNS RP Years 9 to 12 participants were eligible for this study (that is refused permission to be re-contacted for future research), see section A.1
- a further adjustment was made to account for differences in the willingness of the contacted NDNS RP Years 9 to 12 participants to complete the web-questionnaire
- the resulting weights were combined and calibrated to ONS mid-year population estimates for 2019 (20) by age group, sex, and region to create the individual weights – this step was repeated for subsequent stages of this study to create the dietary recall and RPAQ weights
- an adjustment for unequal selection probabilities of individuals within households to generate a household-level weight that was subsequently calibrated to household population estimates by region and tertiles of the Index of Multiple Deprivation (IMD)

A.6.1 Starter weights

The survey was designed as a follow up to the NDNS RP Years 9 to 12. As such, the weighting scheme is designed to incorporate the NDNS RP weights to account for differential selection

probabilities of the individuals taking part in NDNS RP, and for non-response bias. The NDNS RP individual weights were used as the starting point of the weighting scheme for this study (21).

For NDNS RP Years 9 to 11 participants the existing, combined weights for Years 9 to 11 were used as the starting weights. However, participants from NDNS RP Year 12 did not yet have an individual weight, hence a set of individual weights to use as starter weights were generated for NDNS RP Year 12 participants following the usual process for NDNS RP weighting. These 2 sets of weights (the NDNS RP Years 9 to 11 individual weights and the NDNS RP Year 12 individual weights) were then combined into a single starter weight.

Finally, the combined starter weights were scaled to so that the weighted sample size matched the unweighted sample size.

A.6.2 Weights for web-completion

A set of weights were required for individuals who had completed the web questionnaire. These weights were generated in a number of stages that combined the NDNS RP starter weights, the weights from 2 non-response models, and a calibration step. These are described below.

A.5.6.1 Model 1

The aim of this first step was to reduce bias caused by differences between those individuals who consented to be re-contacted and were subsequently invited to participate in this study, and those who did not. There were also 69 NDNS Year 12 cases that were eligible for this study but were not included in the issued sample as their NDNS RP data was still being processed and cleaned at the time the sample was drawn. These cases were included as non-respondents in the logistic regression model, meaning the model adjusts the sample to correct for any coverage error as well as non-response. Individuals from NDNS RP Years 9 to 11 who had completed 3 or 4 food diary days, and individuals from NDNS RP Year 12 who had completed at least one dietary recall day, who had consented to be contacted for follow-up research were eligible for this study.

A logistic regression model was run using information about the individual taken from the NDNS RP Years 9 to 12 interview, namely, age of respondent at time of sampling for this study, sex, region, ethnicity, household size, number of children in the household, tenure, whether the HRP was in work, population density of the local area (22) and IMD quintiles (23). NDNS RP survey year was also included in the model. Only those variables significantly related to the outcome were retained in the final model. The model used a stepwise procedure to select those variables that were most strongly associated with the likelihood that an individual was issued for this study. Any variables that were not significantly related to response behaviour were dropped. The non-response model was run on data weighted by the starter weights to ensure the model was summarising patterns of non-response around re-contact, rather than differences in sample

composition that have already been corrected for by the NDNS RP weights, such as unequal selection probabilities.

The model summarises the non-response behaviour of the individuals and was used to generate, for each individual, the predicted probability that they were issued for this study, given their individual, household, and area characteristics. Weights were generated as the inverse of the predicted probability of being issued. These weights were checked for outliers and any large weights were trimmed (24). The weights were then combined with the NDNS RP starter weights.

This step therefore adjusts the sample to address bias caused by not having completed sufficient diary or dietary recall days in their previous NDNS RP participation to be eligible and refusing permission to be re-contacted for future research.

A.6.2.2 Model 2

A second model was run to model non-response to this study by the issued sample. The same set of variables from NDNS RP were used in the modelling and the same stepwise process was used. The data were weighted by a weight that combined the NDNS RP starter weight and the weight from model one to account for bias preceding this step. As before, a set of weights were generated as the inverse of the predicted probabilities of response produced by the model. These weights were checked for outliers and trimmed, then combined with the starter weights and the weights from model 1.

A.6.2.3 Calibration step

The next step was to calibrate the individual weights created thus far for participants completing the web questionnaire to ONS 2019 mid-year population estimates (20) for age, sex and region. The calibration was run separately for adults (aged 19 years and over) and children (aged 1.5 to 18 years) to ensure the regional distribution was correct within each group. These calibrated weights are the final weights for the web questionnaire participants. The distributions of the population and weighted and unweighted achieved samples are shown in table A.6.1 for each of the 3 individual weights.

A.6.3 Dietary recall weights

Not all individuals who completed the web questionnaire went on to complete any dietary recalls. Thus an additional weighting stage was needed for the dietary recall weight (for participants who completed at least one recall) to reduce non-response bias caused by participants dropping out. A second calibration step was run on this sub-set of respondents, calibrating to the same set of ONS population figures and using the combined model and NDNS RP weights as a set of starter weights. This calibrated weight is the final dietary weight.

A.6.4 Physical activity (RPAQ) weight

Not all individuals were eligible for RPAQ, only those aged 16 years or over who had completed all 4 dietary recalls, and who had provided an email address were invited to take part. The RPAQ weight was generated by calibrating the individual weights of the RPAQ respondents to ONS 2019 mid-year population estimates (20) for age and sex, and region, for individuals aged 16 years and over. These weights are the final RPAQ weights.

A.6.5 Household weights

As described in section A.5, all household-level data from children aged 11 to 15 years were excluded from household-level analysis and reporting. Household weights were only generated for households which contained a participant outside this age range.

NDNS RP is designed to be an individual-level survey and does not have a household weight. This study followed up individuals, rather than households, hence it is also an individual-level survey. This meant there was no household-level NDNS RP weight to use as a starter weight, and therefore the household weights are created by amending the NDNS RP individual weights.

The first step in the household weighting was to reverse the adjustment for individual selection that was carried out as part of the NDNS RP weighting, thus converting the weights from individual-level to household-level. This adjustment was made using the information collected in NDNS RP Years 9 to 12 that was used to generate the individual selection weights. These reverse-selection weights are applied to the individual weights, effectively 'undoing' the individual selection for each responding individual. These adjusted weights were then aggregated to household level. Where there was more than one participant per household, the mean weight was taken.

These weights were then calibrated to population estimates of the regional distribution of private households, taken from the Labour Force Survey (LFS) (25) and to tertiles based on the IMD. This calibrated weight is the final household weight. The distributions of the population and weighted and unweighted achieved sample are shown in table A.6.2 for the household weights.

A.7 Statistical analysis

A.7.1 Participant-matched statistical comparison of dietary intake and physical activity between this study and previous NDNS RP assessments

This section outlines the statistical methods used to estimate the change in dietary intake and physical activity from participants' previous NDNS RP assessment (obtained between April 2016 and March 2020) and their assessment for this study (obtained between August and October 2020). Each participant was surveyed during both time periods, so a participant

matched comparison was analysed by calculating the change in dietary intakes or physical activity energy expenditure (PAEE) from the 2 time-periods.

Linear and logistic regression models were used to estimate the average change with 95% confidence interval for each age and sex group (2 to 10 years, 11 to 18 years, 19 to 64 years, 65 years and over for males, females and sex combined). These regression models included covariates to adjust for seasonality (month of completion) and the 'time between assessments' (0.5-4.5 years). All analyses took into account the complex survey design.

Linear regression models were used for continuous measurements. The regression coefficients in this model were estimated using probability weighted least squares (26) and their covariance matrix was estimated using a Taylor linearization method (27).

Logistic regression models (with an identity link function) were used for binary variables. The regression coefficients (which estimate the proportion parameters for each age or sex group) use a pseudo-likelihood approach (26) and their covariance matrix was estimated using a Taylor linearization method (27). The proportion parameter (along with the associated 95% confidence interval) estimates the proportion of people above the threshold for each variable.

An adjustment was made to participants' previous NDNS RP dietary intake data to account for the natural age progression expected for each variable between the assessments. Data from the NDNS RP Years 1 to 11 (2008/09 to 2018/19) was analysed using a segmented regression method to inform the required adjustment with age for respective variables. Table A.2 shows the adjustment made for each variable.

Table A.2 Adjustments made to previous NDNS RP dietary intake data to account for the natural age progression

Variable	Sex	Per year adjustment before age cut point	Age at cut point (years)	Per year adjustment after age cut point
Total meat (g/day)	Male	+5.81	20.0	-0.86
	Female	+2.86	18.6	-0.38
Red and processed meat (g/day)	Male	+3.56	19.1	-0.25
	Female	+2.12	10.8	+0.055
Total fish (g/day)	Male	+0.31	69.0	-0.21
	Female	+0.28	71.0	-0.35
Soft drinks not low calorie (g/day)	Male	+17.53	16.0	-4.50
	Female	+12.29	14.9	-3.06
Chocolate confectionery (g/day)	Male	+0.78	13.8	-0.15
	Female	+0.65	11.6	-0.10

Biscuits (g/day)	**Male**	+1.17	5.9	-0.039
	Female	+1.95	4.4	-0.0063
Buns, cakes and pastries (g/day)	**Male**	+3.03	4.3	+0.032
	Female	+2.98	5.0	+0.027
Crisps and savoury snacks (g/day)	**Male**	+0.73	12.8	-0.18
	Female	+1.06	9.0	-0.16
Total energy (kcal/day)	**Male**	+57.27	20.3	-7.17
	Female	+72.44	9.5	-1.49
Total energy (MJ/day)	**Male**	+0.24	20.3	-0.030
	Female	+0.30	9.5	-0.006
Carbohydrate (g/day)	**Male**	+8.81	14.3	-0.80
	Female	+12.20	8.0	-0.51
Free sugars (g/day)	**Male**	+3.62	14.0	-0.50
	Female	+4.04	9.3	-0.40
Total fat (g/day)	**Male**	+2.18	19.5	-0.24
	Female	+2.59	9.5	-0.039
Saturated fatty acids (g/day)	**Male**	+0.76	15.4	-0.025
	Female	+0.86	7.0	+0.012
Protein (g/day)	**Male**	+2.36	22.0	-0.31
	Female	+0.88	27.2	-0.12
Fibre (g/day)	**Male**	+0.35	25.4	-0.022
	Female	+0.17	37.6	-0.028
Folate (mg/day)	**Male**	+5.04	26.3	-0.19
	Female	+1.35	61.7	-1.88
Iron (mg/day)	**Male**	+0.35	15.3	-0.002
	Female	+0.43	8.5	+0.011
Sodium (mg/day)	**Male**	+83.0	19.3	-10.5
	Female	+76.1	11.5	-3.2
Alcohol (g/day)	**Male**	+0.63	47.7	-0.37
	Female	+0.43	36.9	-0.16
Alcohol – consumers only (g/day)	**Male**	+0.55	47.5	-0.48
	Female	+0.34	37.0	-0.22

For example, if a male participant was aged 11 years for their previous NDNS RP dietary assessment and aged 14 years for this study, their previous NDNS RP total meat (g per day) value would be increased by 3*5.81=17.43 g per day (figure A.1). If a male participant was aged 19 years for their previous NDNS RP dietary assessment and aged 23 years for this study, their previous NDNS RP total meat (g per day) value would be increased by 1*5.81=5.81 g per day

and decreased by 3*0.86=2.58 g per day resulting in a net increase of 3.23 g per day. If the adjustment resulted in a negative value for a participant this was truncated to zero.

Figure A.1
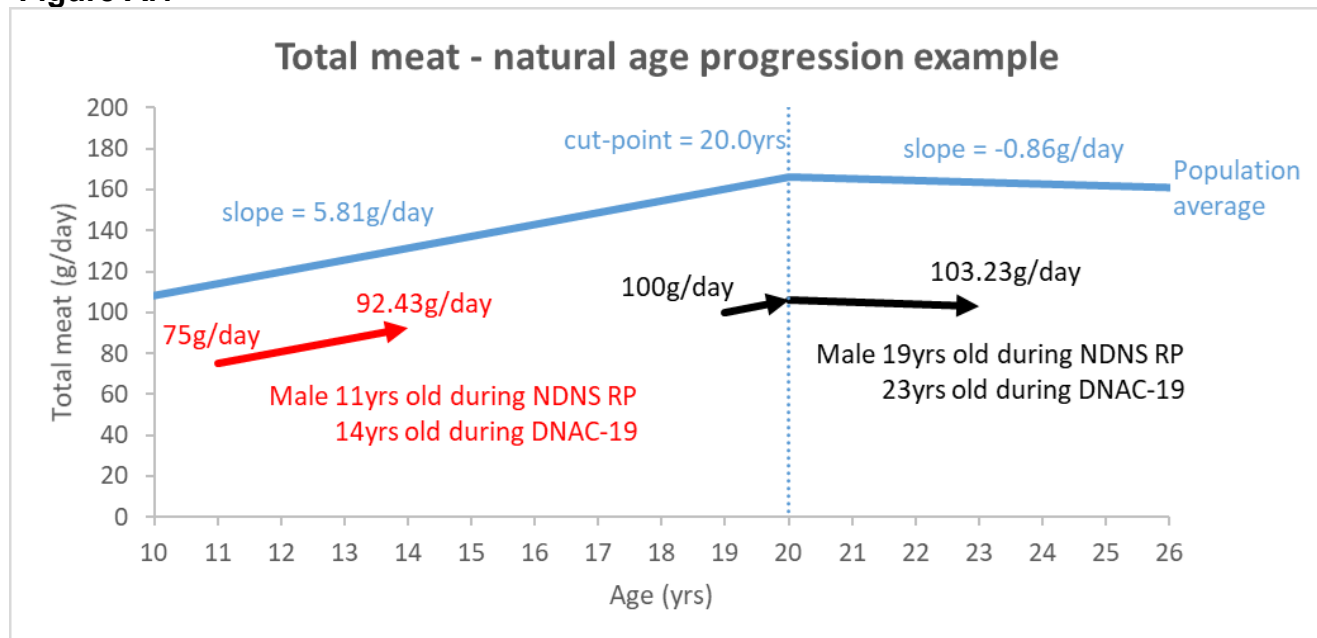
It was noted that the age at cut point was very low for some variables (for example Biscuits; males = 5.9 years, females = 4.4 years). Further investigations found that this was partly due to the percentage of consumers not being consistent with age (for example of those aged 10 years 80% were biscuit consumers whereas of those aged 20 years only 50% were biscuit consumers – but the relationship with age was not linear) and partly due to the foods included within the variable (for example, Biscuits includes both sweet and savoury foods which are consumed differently for different ages). When the segmented regression method was applied to consumers only of Biscuits the age at cut point increased (males = 15.0 years, females = 10.2 years). However, the natural age progression adjustment required for this participant-matched comparison is based on the whole population (including non-consumers) so the figures in table A.2 were used to make adjustments to participants' previous NDNS RP dietary intake data.

No adjustments were made for total fruit and vegetable portions, sugar confectionery, vitamin D or calcium because the regression analysis showed no change with age.

The following NDNS RP variables were recalculated using the adjusted variables before being analysed: Food energy (kcal per day), Food energy (MJ per day), Carbohydrate (% food energy and % total energy), Free sugars (% food energy and % total energy), Total fat (% food energy and % total energy), Saturated fatty acids (% food energy and % total energy), Protein (% food energy and % total energy), Alcohol (% total energy), Free sugars below threshold indicator, Fibre below threshold indicator, Saturated fatty acids below threshold indicator.

The goodness of fit of the linear models was examined using the concept of explained variation (R-squared).

The statistical analyses described above were performed using the survey package in the statistical program R (28, 29, 30).

The statistical analyses described in this appendix are for descriptive purposes rather than analytical, that is, they are not intended to estimate the associations among many variables. Therefore, corrections for multiple comparisons were not necessary. Bonferroni procedures may be applicable in other situations involving simultaneous testing of regression coefficients when the number of independent variables in the regression analysis is large compared to the number of sampled Primary Sampling Units (PSUs) (31).

References

1. 'Eligible participants' were defined as those who had completed at least 3 diary days (Years 9-11) or 1 diet recall day (Year 12) and agreed to be re-contacted about future research
2. The change in household sampling model in Year 12 is considered as part of an ongoing evaluation of changes made within the NDNS RP. Further details can be found in 'Evaluation of change in dietary methodology in NDNS rolling programme: Stage 1'
3. 'National Diet and Nutrition Survey: Results from Years 9 to 11 (combined) 2016/2017 to 2018/2019' (viewed on 13 August 2021)
4. 'Evaluation of change in dietary methodology in NDNS rolling programme: Stage 1'
5. Fitt E and others. 'DINO (Diet In Nutrients Out) – an integrated dietary assessment system' Public Health Nutrition 2015: volume 18 issue 2, pages 234-41 (viewed 16 August 2021)
6. Rowland MK and others. 'Field Testing of the Use of Intake24 – An Online 24-Hour Dietary Recall System' Nutrients 2018: volume 10, page 1,690 (viewed 16 August 2021)
7. Bradley J and others. 'Comparison of INTAKE24 (an Online 24-h Dietary Recall Tool) with Interviewer-Led 24-h Recall in 11-24 Year-Old' Nutrients 2016: volume 8 issue 6, page 358 (viewed 16 August 2021)
8. Foster E and others. 'Validity and reliability of an online self-report 24-h dietary recall method (Intake24): a doubly labelled water study and repeated-measures analysis' Journal of Nutritional Science 2019: volume 8, e29 (viewed 16 August 2021)
9. The demo version of Intake24 is updated regularly and so may not be exactly the same as the version used in this study.
10. The shielding programme was paused on 31 July for Northern Ireland, 1 August for England and Scotland and 16 August for Wales and so was not happening over the dietary data collection period.
11. Main and subsidiary NDNS food groups are listed in appendix R of the 'National Diet and Nutrition Survey Years 1 to 9 of the Rolling Programme (2008/2009 – 2016/2017): Time trend and income analyses' (viewed on 13 August 2021)
12. During the early stages of Year 12 data collection, extreme portions of pizza were identified. Adjustments were carried out at the food group level by applying winsorization which involves recoding extreme values to the nearest 'reasonable' value (Data processing). A review of the portion size pathway in Intake24 indicated an issue with the selection of number of 'slices' versus 'whole' pizza. The pathway was immediately amended in the tool to address this.
13. Physical activity downloads
14. Besson H and others. 'Estimating physical activity energy expenditure, sedentary time, and physical activity intensity by self-report in adults' American Journal of Clinical Nutrition 2010: volume 91 issue 1, pages 106-14 (viewed 16 August 2021)
15. Golubic R and others. 'Validity of Electronically Administered Recent Physical Activity Questionnaire (RPAQ) in Ten European Countries' PLoS ONE 2014: volume 9 issue 3, e92829 (viewed 16 August 2021)
16. Ainsworth BE and others. 'Compendium of Physical Activities: a second update of codes and MET values' Medicine and Science in Sports and Exercise 2011: volume 43 issue 8, pages 1,575-81 (viewed 16 August 2021)

17. Ainsworth BE and others. 'The Compendium of Physical Activities Tracking Guide'. Healthy Lifestyles Research Center, College of Nursing and Health Innovation, Arizona State University (viewed 16 August 2021)

18. Lindsay T and others. 'Descriptive epidemiology of physical activity energy expenditure in UK adults (The Fenland study)' International Journal of Behavorial Nutrition and Physical Activity 2019. 9;16(1):126 (viewed 12 August 2021)

19. A hundred and one households where the only participant was aged 11 to 15 years were excluded from household-level analysis and were not included in the household weighting. See appendix A, section A.6.

20. Office for National Statistics (2021). 'Population estimates' (viewed on 13 August 2021)

21. More information about the design of these weights can be round in the appendices of the 'National Diet and Nutrition Survey: Results from Years 9 to 11 (combined) 2016/2017 – 2018/2019' (viewed on 13 August 2021)

22. Measured as population in private households in the local area, divided by hectares, entered into the model as quintiles.

23. The most recent IMD for each country was used.

24. Large weights can inflate the standard errors which has a detrimental impact on sample efficiency. Trimming, or capping, involves reducing the size of the largest weights by replacing the largest weights with a lower value. In this instance, all weights larger than the 99.5th percentile were given the value of the 99.5th weight.

25. Office for National Statistics (2020). 'Labour Force Survey' (viewed on 21 May 2021)

26. Holt D and others. 'Regression analysis of data from complex surveys' Journal of the Royal Statistical Society A 1980: volume 143, pages 474-487

27. Binder DA. 'On the Variances of Asymptotically Normal Estimators from Complex Surveys' International Statistical Review 1983: volume 51, pages 279-292

28. R Core Team (2013). 'R: A language and environment for statistical computing' R Foundation for Statistical Computing, Vienna, Austria

29. Lumley T. 'Analysis of complex survey samples' Journal of Statistical Software 2004: volume 9 issue 1, pages 1-19

30. Lumley T. (2012) 'Survey: Analysis of Complex Survey Samples' R package version 3.28-2.

31. Korn EL and Graubard BI. 'Simultaneous testing of regression coefficients with complex survey data: Use of Bonferroni t statistics' The American Statistician 1990: volume 44, pages 270-276